

Publication number: JP2002244842

Inventor: IFUKUBE TATSU

Applicant: JAPAN SCIENCE & TECH CORP; BUG INC; IFUKUBE
TATSU

Classification:

- international: **G06K9/00; G06K9/00; (IPC1-7): G06F3/16; G06F3/00; G06F17/27; G06F17/28; G10L15/00; G10L15/22**

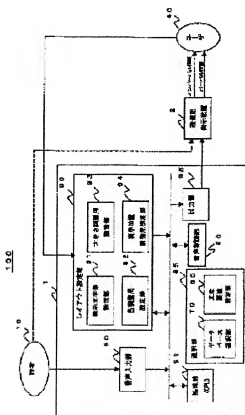
- european: G06K9/00G

Application number: JP20010044794 20010221

Priority number(s): JP20010044794 20010221

[Report a data error here](#)

PROBLEM TO BE SOLVED: To support communications in different language systems between a speaker and a user by providing the user with language information and non-language information from the speaker simultaneously. **SOLUTION:** A computing unit 1 recognizes voice inputted by the speaker 10 through a microphone or the like at a voice recognition part 60 and converts the same into language system which the user 40 has acquired (interpretation) at an interpretation part 85 to provide language information (verbal information, i.e., character string). The character string as a result of the voice recognition and interpretation is outputted on the permeable display device 2. The permeable display device 2 indicates the character string of language information inputted by the computing device 1 and is provided with a permeable part (permeable display) for providing non-language information (non-verbal information, for example, notions of parts of the speaker's face such as lip and eyes, reading, gesture, sign language, expression or some of those) from the speaker 10.



<http://v3.espacenet.com/textdoc?DB=EPODOC&IDX=JP2002244842&F=0>

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2002-244842
(P2002-244842A)

(43) 公開日 平成14年8月30日 (2002.8.30)

(51) Int.Cl. ⁷	識別番号	F I	キーワード ⁷ (参考)
G 0 6 F 3/16	3 4 0	C 0 6 F 3/16	3 4 0 F 5 B 0 9 1
	3 2 0		3 2 0 D 5 D 0 1 5
	3/00	3/00	6 0 1 5 E 5 0 1
	17/27	17/27	J
	17/28	17/28	T

審査請求 未請求 請求項の数 9 O L (全 10 頁) 最終頁に続く

(21) 出願番号 特願2001-44794(P2001-44794)

(22) 出願日 平成13年2月21日 (2001.2.21)

特許法第30条第1項適用申請有り

(71) 出願人 398020800

科学技術振興事業団

埼玉県川口市本町4丁目1番8号

(71) 出願人 591271450

株式会社ビー・ユー・シー

北海道札幌市厚別区下野幌テクノパーク1
丁目1番14号

(71) 出願人 599008654

伊福部 達

北海道札幌市中央区南13条西13丁目1-43

(74) 代理人 100107010

弁理士 橋爪 健

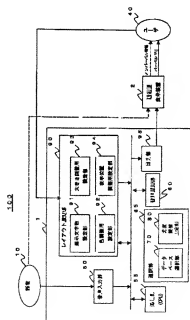
最終頁に続く

(54) 【発明の名称】 音声通訳システム及び音声通訳プログラム

(57) 【要約】 (修正有)

【課題】 話者による言語情報と非言語情報とを、同時にユーザに显示することで、話者とユーザとの異なる言語体系間におけるコミュニケーションの補助を行う。

【解決手段】 演算装置1は、音声認識部60で話者10によりマイクロホン等を介して入力された音声进行認識し、さらに、通訳部85でユーザ40の獲得している言語体系に変換(通訳)することにより、言語情報(バーバル情報)であって、ここでは、文字列を得る。この音声認識及び通訳の結果である文字列を透過型表示装置2に出力する。透過型表示装置2は、演算装置1により入力された言語情報である文字列を表示すると共に、話者10からの非言語情報(ノンバーバル情報)であって、例えば、話者10の唇、目等の顔の部分の動き、読話、ジェスチャー、手話、表情のいずれか又は複数を含む)を得るための透過部(透過性ディスプレイ)を備える。



【特許請求の範囲】

【請求項1】異なる言語体系間のコミュニケーションを補助するための音声表示システムであって、入力された音声認識して言語情報に変換して出力する演算装置と、前記言語情報を表示すると共に、話者からの非言語情報を得るための透過部を含む透過型表示装置とを備え、前記演算装置は、該音声を入力する音声入力部と、該音声入力部から入力された音声認識するための音声認識部と、前記音声認識部で音声認識された言語情報を、異なる言語体系に翻訳する翻訳部と、前記透過型表示装置への表示状態を設定するためのレイアウト設定部と、前記レイアウト設定部の設定に従って、前記音声認識部により音声認識され、且つ、前記翻訳部で翻訳された結果を、前記透過型表示装置に出力する出力部とを有する音声通訳システム。

【請求項2】前記レイアウト設定部は、前記透過型表示装置に表示される文字列の表示文字数、色、大きさ、表示位置、表示の焦点距離のいずれか又は複数を設定するようにした請求項1に記載の音声通訳システム。

【請求項3】前記非言語情報は、該話者の唇、目の顔の部分の動き、ジェスチャー、手話、読話、表情のいずれか又は複数を含むことを特徴とする請求項1又は2に記載の音声通訳システム。

【請求項4】前記レイアウト設定部では、該読話又は手話の習熟度が高い場合、前記表示文字数又はその割合を小さく又は大きくし、

該習熟度が低い場合、前記表示文字数又はその割合を大きく又は小さくそれぞれ設定されるようにした請求項2又は3に記載の音声通訳システム。

【請求項5】前記演算装置は、携帯型のコンピュータであって、前記透過型表示装置は、メガネディスプレイであることを特徴とする請求項1乃至4のいずれかに記載の音声通訳システム。

【請求項6】前記通訳部は、設定された漢字の難易度に応じて選択された言語データベースに基づき、前記表示される文字列を漢字に変換するようにした請求項1乃至5のいずれかに記載の音声通訳システム。

【請求項7】前記通訳部は、音声認識された文字列の尤度の閾値を設定する尤度閾値設定部を含み、該尤度閾値設定部で予め設定された閾値に応じて、尤度が該閾値より低いときは、前記表示される文字列を非文字で表示するようにした請求項1乃至6のいずれかに記載の音声通訳システム。

【請求項8】入力された音声認識して言語情報に変換して出力する演算装置と、前記言語情報を表示すると共に

に話者からの非言語情報を得るための透過部を含む透過型表示装置とを備えた異なる言語体系間のコミュニケーションを補助するための音声表示システムに用いられる音声表示プログラムであって、該音声を入力する音声入力部と、音声認識された文字列の尤度についての予め設定された閾値に応じて、該尤度が該閾値より低いときは、表示される文字列を非文字で表示するように認識するようにした、該音声入力部から入力された音声認識するための音声認識部と、前記音声認識部により音声認識された言語情報を、異なる言語体系に翻訳する翻訳部と、前記音声認識部により音声認識され、且つ、前記翻訳部により翻訳された結果を、前記透過型表示装置に出力する出力部とをコンピュータに実行させるための音声通訳プログラム。

【請求項9】前記通訳部は、設定された漢字の難易度に応じて選択された言語データベースに基づき、前記表示される文字列を漢字に変換するようにした請求項8に記載の音声通訳プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声通訳システム及び音声通訳プログラムに係り、特に、音声認識結果データである文字列の情報（言語情報：バーバル情報）だけでなく、話し手（話者）の表情、唇、ジェスチャー等の言語情報以外の情報（非言語情報：ノンバーバル情報）を用いて、話者とユーザとの異なる言語体系間におけるコミュニケーションの補助を行うことができる音声通訳システム及び音声通訳プログラムに関する。

【0002】

【従来の技術】近年、高度情報化および超高齢化の社会においては、各種情報を受け取る感覚器官（例えば、視覚、聴覚等）の能力が低い人達（例えば、高齢者、聴覚障害者等）のために、これらの感覚器官の能力を補うための各種補助装置（例えば、補聴器等）の開発が切望されている。特に、人間同士のコミュニケーションでは、音声非常に重要な役割を果たしており、聴覚障害者のための様々な補助方式が研究されている。例えば、大学の講義において、講義内容を素早く書き取り、話者又は講義用黒板の周辺に字幕を表示する手法がある（参考：小林正幸・石原保志・西川俊・高橋秀紀、ルビ付きリアルタイム字幕提示システムの試作；筑波技術短期大学テクノレポート、1996）。

【0003】

【発明が解決しようとする課題】しかしながら、上述の手法では、第三者の助けが必要であり、例えば、聴覚障害者が単独で外出する場合での利用は困難であることが想定される。このように、未だに聴覚障害者が日常的に利用することができる補助装置は存在していない。した

がって、聴覚障害者のための補助装置は、今後ますます増え続けるであろう高齢聴覚者、又は完全聴覚障害者にとっても有用であり、その必要性は今後も高まると考えられる。

【0004】一方、近年、入力された音声認識して文字列に変換し、この文字列を、例えば、コンピュータのモニタディスプレイやテレビの字幕として表示する、いわゆる音声認識方法が普及している。この音声認識方法は、雑音の多い環境（すなわち日常生活空間）で不特定話者を対象に認識を行った場合、一般には、50～60%程度の認識率しか確保できないため、特定の用途での利用に限られているのが現状である。

【0005】しかし、この現状は、音声認識システムを、従来のようにマン・マシン・インターフェースとして捉えた結果であり、この音声認識システムを、人間を対象としたマン・マン・インターフェースとして捉えた場合、人間は、入力音声の完全な音声認識が行われなくてもコミュニケーションにおける前後の文脈などから欠落した情報を類推することができる（参照：齊藤幹、失聴者のための音声認識技術を利用したマン・マン・インターフェースに関する研究；北海道大学大学院工学研究科修士論文、1999年2）。

【0006】ここで、本発明に関連する技術について説明する。本発明者らは、コミュニケーションでは、話者の音声以外にも唇、目等の顔の動き、談話（唇の動きを読むこと）、ジェスチャー、手話、表情といった非言語情報（ノンバーバル情報）も重要であり、さらに、このノンバーバル情報は、マン・マシン・インターフェースで利用するには非常に高度な技術を必要とするが、マン・マン・インターフェースであれば、人間の視覚を用いることで、容易に取得できる点に着目した。

【0007】また、聴覚障害者（ユーザ）は、上述の「談話」やジェスチャーで伝ええる「手話」を習得している場合が多く、この談話、手話等によって、話者の言葉にある程度理解できることが想定される。このため、音声認識結果の文字列（バーバル情報）と、談話や手話から得られるノンバーバル情報とを、聴覚障害者が同時に受け取れるようにして、音声認識結果の文字列を聴覚障害者に示示することが必要となる。

【0008】この際、考慮すべき点は、談話や手話で文意を理解する能力は、聴覚障害者の失聴時期、残存聴力などに大きく依存している点と、たとえ高性能の音声認識装置による音声認識結果であっても認識率が必ず100%であるとはなく、大抵の場合、文字列には誤りが含まれているため、音声認識結果である文字を全て呈示してしまうと、文意を誤って理解する場合が想定される点である。このため、例えば、聴覚障害者の談話や手話による文理解能力（すなわち、談話や手話の習熟度）に応じて、示示する文字列を表示する割合を、聴覚障害者自身で設定可能とすることが必要である。

【0009】さらに、話者が、ユーザとのコミュニケーションにおいて、ユーザの獲得している言語体系（例えば、母国語）とは異なる言語体系（例えば、外国語）を用いた場合、話者とユーザ間のコミュニケーションを円滑に行うには、話者による言語情報を、ユーザ（受け手）の獲得している言語体系やこの言語体系に基づいた談話や手話などの文理解能力で理解できる言語情報に変換（通訳）することが必要となる。

【0010】本発明は、以上の点に鑑み、話者による言語情報と非言語情報とを、同時にユーザに呈示することで、異なる言語体系間の話者とユーザとのコミュニケーションの補助を行うことを目的とする。また、本発明は、ユーザの特性と音声認識装置の性能に応じて、談話、手話による非言語情報と、音声認識及び通訳結果である言語情報とを、同時にユーザに呈示することを目的としている。また、本発明は、特に、談話や手話による文理解能力が低い中絶失聴者であっても、円滑なコミュニケーションを図ることができることを目的とする。

【0011】

【課題を解決するための手段】本発明の第1の解決手段によると、異なる言語体系間のコミュニケーションを補助するための音声表示システムであって、入力された音声認識して言語情報に変換して出力する演算装置と、前記言語情報を表示すると共に、話者からの非言語情報を得るための透過部を含む透過型表示装置とを備え、前記演算装置は、該音声を入力する音声入力部と、該音声入力部から入力された音声を確認するための音声認識部と、前記音声認識部で音声認識された言語情報を、異なる言語体系に通訳する通訳部と、前記透過型表示装置への表示状態を設定するためのレイアウト設定部と、前記レイアウト設定部の設定に従って、前記音声認識部により音声認識され、且つ、前記通訳部で通訳された結果を、前記透過型表示装置に出力する出力部とを有する音声通訳システムを提供する。

【0012】本発明の第2の解決手段によると、入力された音声認識して言語情報に変換して出力する演算装置と、前記言語情報を表示すると共に話者からの非言語情報を得るための透過部を含む透過型表示装置とを備えた異なる言語体系間のコミュニケーションを補助するための音声表示システムに用いられる音声表示プログラムであって、該音声を入力する音声入力手順と、音声認識された文字列の尤度についての予め設定された閾値に応じて、該尤度が該閾値より低いときは、表示される文字列を非文字で表示するように認識するようにした、該音声入力手順から入力された音声を確認するための音声認識手順と、前記音声認識手順により音声認識された言語情報を、異なる言語体系に通訳する通訳手順と、前記音声認識手順により音声認識され、且つ、前記通訳手順により通訳された結果を、前記透過型表示装置に出力する出力手順とをコンピュータに実行させるための音声通訳プ

ログラムを提供する。

【0013】また、本発明は、誤りを含む音声認識及び通訳結果データである言語情報（文字列）を、透過型表示装置（メガネディスプレイ）上に表示することにより、この透過型メガネディスプレイを用いるユーザは、話者の前に表示された文字列だけでなく、話者の唇、目等の顔の部分の動き、読話、ジェスチャー、手話、表情などを含む非言語情報をも同時に見ることができ、これにより、ユーザが聴覚障害者であっても、話者の文意を理解しやすくなり、異なる言語体系間におけるユーザと話者間の円滑なコミュニケーションを図ることができる。

【0014】

【発明の実施の形態】以下、図面を用いて本発明の実施の形態を詳細に説明する。図1は、本発明に関する音声通訳システム100の概略構成図である。音声通訳システム100は、例えば、携帯型のコンピュータである演算装置1、透過型メガネディスプレイである透過型表示装置2を含む。この音声通訳システム100は、話者10とユーザ（例えば、聴覚障害者など）40間の異なる言語体系間におけるコミュニケーションを補助するためのシステムである。なお、異なる言語体系間におけるコミュニケーションとは、例えば、話者10がユーザ40の獲得している言語体系（例えば、母国語）とは異なる言語体系（例えば、外国語）を用いて、ユーザ40とのコミュニケーションを行う場合を想定している。

【0015】演算装置1は、話者10により図示しないマイクホン等を介して入力された音声进行認識し、さらにユーザ40の獲得している言語体系に変換（通訳）することにより、言語情報（パナール情報であって、ここでは、文字列）を得る。この音声認識及び通訳の結果である文字列を透過型表示装置2に出力する。透過型表示装置2は、演算装置1により入力された言語情報である文字列を表示すると共に、話者10からの非言語情報（ノンパナール情報であって、例えば、話者10の唇、目等の顔の部分の動き、読話、ジェスチャー、手話、表情のいずれか又は複数を含む：図中、点線の矢印）を得るための透過型（透過性ディスプレイ）を備える。

【0016】演算装置1は、例えば、音声入力部50、処理部（CPU）55、音声認識部60、通訳部85、レイアウト設定部90及び出力部95を備える。音声入力部50は、話者10の音声を入力する。音声認識部60は、例えば、音声入力部50から入力された音声の音声認識を行うものである。通訳部85は、例えば、音声認識部60で音声認識された言語情報を、ユーザ40の獲得している言語体系に通訳するものであって、データベース選択部70と、尤度閾値設定部80を備える。具体的には、通訳部85は、例えば、話者10がユーザ40の獲得している言語体系（例えば、母国語）とは異なる言語体系（例えば、外国語）を用いて、ユーザ40と

のコミュニケーションを行う場合、音声認識部60より音声認識された話者10の言語体系による言語情報を、ユーザ40の獲得している言語体系に変換（通訳）を行うものである。

【0017】データベース選択部70は、例えば、通訳部85内又は音声通訳システム100内に適宜設けられたひとつ又は複数種類の言語データベース（漢字DB）に予め記憶された漢字の難易度（例えば、JIS水準レベルに従って、小学2年生レベル、中学生レベルなど）を、ユーザ40の語彙力（どの程度の漢字を含む文字列を作成するべきかを判断することになる）に応じて選択する。これにより、話者10の音声进行認識する際、ユーザ40の語彙力に応じた漢字を含む文字列を作成することができる。なお、言語データベースにひらがな該当する漢字を認識しない場合、文字列は、全てひらがな及び／又は非文字（例えば、記号など）として透過型表示装置2の透過性ディスプレイ上に表示される。

【0018】また、通訳部85は、図示しない音声データベースに記憶された話者10からの連続音声に対して、言語データベースを用いて音声認識が行われた場合、単語系列と各単語の尤度（例えば、単語間のつながりの確からしさであって、単語間の距離と捉えよう良い。したがって、尤度が高いことと、単語間の距離が小さくことは、概ね同義である）を認識結果として得る。

【0019】尤度閾値設定部80は、例えば、通訳部85での認識結果のうち正しく認識できた文字列のみを呈示するために尤度に対応した閾値を設定する。具体的には、各単語の尤度が高い場合、その単語が正しく認識できている可能性が高く、尤度が低い場合、その単語が正しく認識できている可能性が低いと想定される。このため、尤度閾値設定部80で閾値を設定することにより、認識結果の単語系列に対して、各単語の対数尤度がある閾値より大きければ、この単語を呈示し、対数尤度が閾値より小さいならば、この単語を呈示しないようにすることもできる（閾値の具体的な数値については、後述）。なお、通訳部85では、この対数尤度が閾値より小さい単語（単語間のつながりの確からしさが小さく、単語間の距離が大きい）を呈示しないだけでなく、記号等の非文字として透過型表示装置2の透過性ディスプレイ上に表示するように適宜設定することができる。

【0020】レイアウト設定部90は、透過型表示装置2への表示状態を設定するものであって、例えば、呈示文字数設定部91、色調整用設定部92、大きさ調整用設定部93及び表示位置調整用設定部94を備える。呈示文字数設定部91は、例えば、演算装置3より出力され、透過型表示装置2に表示される文字列の呈示文字数を適宜調整することができる。また、呈示文字数設定部91は、例えば、ユーザ40の読話及び／又は手話の習熟度と呈示文字数との対応を示す習熟度フォルダを含む

(図2参照)。

【0021】色調整用設定部92は、同じく、文字列の色を適宜調整することができる。大きさ調整用設定部93は、同じく、文字列の大きさを適宜調整することができる。表示位置調整用設定部94は、同じく、文字列の表示位置、表示の焦点距離(例えば、話者10の距離に応じて、文字列の表示される焦点距離を調整可能)を適宜調整することができる。なお、レイアウト設定部90に含まれる各種設定部の設定は、ユーザ40自身により適宜設定される(図中、実線の矢印)。これにより、透過型表示装置2に表示される文字列のレイアウトは、ユーザ40の所望する状態に変更することができる。出力部95は、ユーザ40によるレイアウト設定部91の設定に従って、音声認識部60により音声認識された結果である文字列(バーバル情報)を、透過表示装置2に出力する。

【0022】図2は、習熟度フォルダ20の説明図である。習熟度フォルダ20は、上述のように、レイアウト設定部90内の表示文字数設定部91に含まれており、例えば、ユーザ40の読話及び/又は手話の習熟度21と表示文字数(の割合)22との対応を示している。ここでは、習熟度21と表示文字数(の割合)22との対応としては、「低い、80%」「普通、60%」「高い、40%」を予め記憶している。なお、この習熟度フォルダ20による習熟度21と表示文字数(の割合)22との対応は、適宜設定することができる。

【0023】ユーザ40(例えば、聴覚障害者)は、透過型表示装置2の透過性ディスプレイにより、話者10の音声(音声認識の結果である文字列(バーバル情報))だけでなく、透過性ディスプレイを介して話者10のノンバーバル情報をも取得することができる。ユーザ40は、例えば、図示しない習熟度設定ボタンなどを用いて、ユーザ40自身の習熟度21を演算装置1に入力する。なお、ユーザ40自身の習熟度21を演算装置1に入力する場合、ユーザ40だけでなく保守者、家族、医者など適宜の人間によって設定するようにしてもよい。

【0024】ユーザ40は、読話、手話についての習熟度21が高い場合(この習熟度は、個人差が大きい)、ノンバーバル情報を読み取ることが、話者10とのコミュニケーションを円滑に行うことが想定される。この場合、ユーザ40にとって文字列は、補助(又は確認)として機能することになる(ここでは、「高い、40%」)。

【0025】一方、ユーザ40は、例えば、読話、手話についての習熟度21が低い場合、ノンバーバル情報だけでは、話者10とのコミュニケーションを行うことが困難であることが想定される。この場合、ユーザ40は、話者10とのコミュニケーションを行うためにバーバル情報に依存する(ここでは、「低い、80%」)。なお、

本発明に関する音声通訳システム100は、聴覚障害者の読話、手話についての習熟度21を向上させるための、一種の訓練システムにも適用できる。具体的には、聴覚障害者の読話、手話についての習熟度21が向上するにつれて、表示文字数(の割合)22を小さくしたり、又は、習熟度21を確認するために表示文字数(の割合)22を大きくしたりしてもよい。

【0026】図3は、本発明に関する音声通訳システム100の使用状態を示す概略説明図である。音声通訳システム100においてユーザ40は、透過型メガネディスプレイ2を装着する。ユーザ40は、透過型メガネディスプレイ2上に表示されるバーバル情報(ここでは、音声認識部60により音声認識され、且つ、通訳部85により通訳された結果データである言語情報「文字列もしくは、・・・など」)だけでなく、透過型メガネディスプレイ2を介して取得される話者10のノンバーバル情報(ここでは、表情、口の動き、ジェスチャー)を用いて話者10とのコミュニケーションを行う。なお、図中、バーバル情報が話者10上に重なるように描かれているが、これは、ユーザ40を主体とすれば、透過型メガネディスプレイ2上に表示される「文字列もしくは、・・・など」は、話者10の手前に表示されているように見えるからである。

【0027】図4は、本発明に関する音声通訳システム100のフローチャートである。まず、表示文字数設定部91は、例えば、図示しない習熟度設定ボタンを介して入力されたユーザ40の習熟度21に関する情報に基づいて、表示文字数(の割合)22を設定する(S201)。なお、ここで、上述の度感閾値の設定及び/又は漢字DBの選択を、必要に応じて行うことができる。つぎに、レイアウト設定部90に含まれる色調整用設定部92、大きさ調整用設定部93及び表示位置調整用設定部94の各種設定を行う(S203)。話者10からの音声(図示しないマイク)を介して音声入力部50に入力される(S205)。

【0028】ステップS205による入力音声(音声認識部60によって、音声認識される(S207))。さらに、ステップS207の音声認識結果データは、通訳部85によってユーザ10の獲得している適宜の言語体系に変換(通訳)される(S213)。ここで、ステップS207及びS213により音声認識され、通訳された言語情報であるバーバル情報は、出力部95を介して、透過型表示装置2に出力される(S209)。透過型表示装置2に出力されるバーバル情報についての変更(ここでは、表示文字数(の割合)22の設定、必要に応じて度感閾値の設定及び/又は漢字DBの選択。さらに、ステップS203による各種設定)が入力されたかを判定する(S211)。ステップS211よりバーバル情報についての変更が入力された場合、再びステップS201及び/又はステップS203に戻り、表示文字

数(の割合)22の設定、さらに、必要に応じて尤度閾値の設定及び/又は漢字DBの選択、さらに、ステップS203による各種設定を行う。また、ステップS21よりバーバル情報についての変更が入力されていない場合、一連の処理を終了する。

【0029】図5は、被験者による文意の理解を客観的に示す実験結果を示す図である。但し、ここでは、従来技術(参照:青森幹、失聴者のための音声認識技術を利用したマン・マン・インターフェースに関する研究;北海道大学大学院工学研究科修士論文、19992)で示された実験を行ったものであるため、本発明の前提条件「人間は、例えば、音声認識結果の文字列の認識率が60%程度確保されれば(不完全なバーバル情報)、前後の文脈及び/又は認識結果から文字列を類推し、結果的にコミュニケーションを図ることができる」を示し、さらに、「この前提条件に加えて、本発明者らによるノンバーバル情報の同時表示という着想によれば、コミュニケーションがさらに円滑に行われる」という根拠を導く程度に簡潔に説明する。

【0030】図5(a)は、対数尤度閾値と表示文の変化を示す図である。この対数尤度閾値と表示文の変化30は、例えば、尤度30、表示文32を含む。音声認識部60により、認識単語及びその尤度が算出される。つぎに、予め設定された閾値より大きい尤度に対応する認識単語のみを表示する。なお、閾値より小さい尤度に対応する単語については、非文字(ここでは、「?」)として表示した。これにより、尤度30と表示文32との対応としては、図示のように、対数尤度閾値を小さくすることで、表示する単語数が増加している。

【0031】図5(b)は、対数尤度閾値による文意理解精度の変化を示す図である。ここでの実験結果としては、図示のように、すべての被験者(A~F)に共通して、対数尤度閾値が「-2500」、すなわち、単語認識精度が約40%を超えるとき文理解精度は急激に上昇している。なお、失聴者においては個人差が大きく、これは獲得語彙数の差が要因のひとつと考えられる。

【0032】以上により、本発明の前提条件「人間は、例えば、音声認識結果の文字列の認識率が60%程度確保されれば(不完全なバーバル情報)、前後の文脈及び/又は認識結果から文字列を類推し、結果的にコミュニケーションを図ることができる」を客観的に示唆した。

【0033】図6は、本発明に関する音声通訳システム100に関連する実験結果を示す図である。本実験は、演算装置1から得られる不完全なバーバル情報と、話者10から得られるノンバーバル情報とを、透過型表示装置2を着用するユーザー40に対して同時表示することによる文章理解変化について実施されたものである。

【0034】本実験では、聴覚障害者に協力してもらう前に、聴覚に障害を持たない23歳~30歳の日本人男性、3名を被験者とした。なお、3名の被験者は、これ

まで特に読話の訓練を受けたことはない。さらに、表示する文章は、図5(a)に示した表示文32と同様とした。

【0035】表示文32と共に表示するノンバーバル情報としては、デジタルビデオ(Victor:GR-DV1、57万画素)で撮影した顔の映像を用いた。この映像は、23歳の日本人男性に音声処理を施す前の正解の文章を読み上げてもらい、その時の顔を中心に撮影した。この映像にPCによるデジタル処理を施し、表示文章字幕を重ね合わせた。文章字幕は、話者が文章をしゃべり終わり口の動きが止まった後に、話者の口元に重ね合わせて表示した。なお、元の文章が同じであれば、4段階の尤度の違い(図6に示すプロット位置に対応)によらず、同じ顔の映像を利用した。

【0036】実験としては、バーバル情報だけの意味理解の変化を調べる第1実験(バーバル情報)と、バーバル情報だけでなく映像試料、すなわちノンバーバル情報を付加した状態での意味理解の変化を調べる第2実験とを行う。第1実験の内容は、図5と同様であり、説明を省略する。また、ここでは、元の文章に応じて無作為に被験者をA、Bの2つのグループに分けた。グループA、Bの被験者には、それぞれ25文に対し、4段階の尤度に分けた100の文章を表示した。各グループの被験者は、例えば、紙に印刷された表示文章を順番に読んでいき、その意味が理解できれば自分の理解した内容を答えるようにした。その際、普段被験者が使っているパソコンを用いて、テキストエディタにキーボードで打ち込ませた。なお、被験者に対しては、指示された順番通りに進み、表示文章を飛ばしたり、前の表示文章に戻ることとはしないように予め教示している。

【0037】第2実験(バーバル情報+ノンバーバル情報)では、映像を被験者に表示するために透過型HMD(OLYMPUS:Mediamark)を用いた。被験者はデジタルビデオを再生し、顔の映像から読話を試み、続いて字幕が現れたところでビデオを一旦停止させ、その文章の内容が理解できれば先ほどと同じようにパソコン上のテキストエディタにキーボードで打ち込ませた。

【0038】また、実験手順としては、第1実験のAグループを試行し、つぎに、第2実験を行い、最後に、第1実験のBグループを試行した。なお、それぞれの実験の間には被験者の判断により休憩を挟んだ。この実験手順を採用したのは、元の文章が同じであり実験を繰り返すことによって生じる文章に対する慣れを少しでも減らすためであり、また、被験者の集中力やモチベーションに結果が影響されやすく、その影響をいくらが少なくするためである。

【0039】また、図示のグラフは、被験者Aの実験結果を示しており、グラフの横軸は対数尤度閾値(単語認識精度)であり、縦軸は文章理解精度(%)である。な

お、四角のマーカ―が施された実線は、第2実験による文理解精度である。また、三角のマーカ―が施された実線は、第1実験でのA、Bグループの結果を平均したものである。

【0040】被験者Aでは、対数尤度閾値が-2000から-3000（認識率で約40%）になると急激に文理解精度が向上している。また、被験者Aで対数尤度閾値が下がれば文理解精度は概ね上昇している。すなわち、被験者Aでは、ノンバーバル情報を付加することによる、文理解精度の向上が明らかに示されている。

【0041】このように本実施の形態の音声通訳システム100によれば、話者による言語情報と非言語情報とを、同時にユーザに呈示することで、話者とユーザとの異なる言語体系におけるコミュニケーションの補助を行うことができる。また、ユーザの特性と音声認識装置の性能に応じて、談話、手話による非言語情報と、音声認識及び通訳結果である言語情報とを、同時にユーザに呈示することができる。また、音声通訳システム100によれば、特に、談話や手話による文理解能力が低い中途失聴者であっても、円滑なコミュニケーションを図ることができる。

【0042】本発明の音声通訳システムは、各部を実現するための機能を含む音声表示方法、その各手順をコンピュータに実行させるための音声通訳プログラム、音声通訳プログラムを記録したコンピュータ読み取り可能な記録媒体、音声通訳プログラムを含みコンピュータの内部メモリにロード可能なプログラム製品、そのプログラムを含むサーバ等のコンピュータ、音声通訳装置、等により提供されることができる。

【0043】

【発明の効果】本発明によると、以上説明した通り、話者による言語情報と非言語情報とを、同時にユーザに呈

示することで、話者とユーザとの異なる言語体系におけるコミュニケーションの補助を行うことができる。また、本発明は、ユーザの特性と音声認識装置の性能に応じて、談話、手話による非言語情報と、音声認識及び通訳結果である言語情報とを、同時にユーザに呈示することができる。また、本発明は、例えば、談話や手話による文理解能力が低い中途失聴者であっても、円滑なコミュニケーションを図ることができる。

【図面の簡単な説明】

【図1】本発明に関する音声通訳システム100の概略構成図。

【図2】習熟度フォルダ20の説明図。

【図3】本発明に関する音声通訳システム100の使用状態を示す概略説明図。

【図4】本発明に関する音声通訳システム100のフローチャート。

【図5】被験者による文意の理解を客観的に示す実験結果を示す図。

【図6】本発明に関する音声通訳システム100に関連する実験結果を示す図。

【符号の説明】

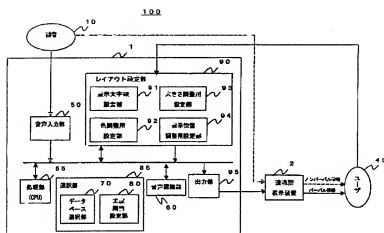
- 1 演算装置
- 2 透過型表示装置
- 10 話者
- 40 ユーザ
- 50 音声入力部
- 55 処理部（CPU）
- 60 音声認識部
- 85 通訳部
- 90 レイアウト設定部
- 95 出力部
- 100 音声表示システム

【図2】

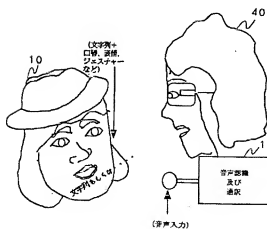
20

21 談話及び／又は手話 の習熟度	22 減文字数の割合
低い	30%
普通	60%
高い	40%

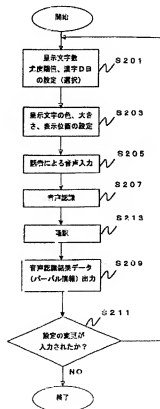
【図1】



【図3】



【図4】



【図5】

30

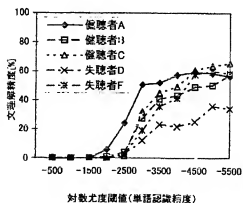
31

32

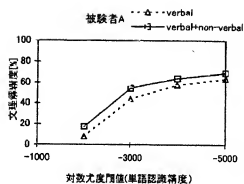
(a)

尤度	呈示文
-3000	??、?を取材した。
-4000	?ばかり、ニューヨークを取材した。
-5000	一週間はかり、ニューヨークを取材した。
-3000	?ばかりが続くと?ないのだ。
-4000	喜びばかりが続くと?ないのだ。
-5000	喜びばかりが続くと分からないのだ。
正解	喜びばかりが続くとは限らないのだ。

(b)



【図6】



フロントページの続き

(51)Int.Cl. ⁷	識別記号	F I	(参考)
G 1 0 L	15/00	G 1 0 L 3/00	5 5 1 C
	15/22		5 6 1 C
(72)発明者	伊福部 達	F ターム(参考)	5B091 AA03 AA15 BA03 BA12 BA19
	北海道札幌市中央区南13条西13丁目1-43		CB12 CB22 EA09
			5D015 HH23 KK03 LL05
			5E501 AA30 BA14 CA06 CB15 CC11
			FA14 FA32